

A Power-Measurement Methodology for Large-Scale, High-Performance Computing

Thomas R. W. Scogland (Virginia Tech & Green500),
Craig P. Steffen (University of Illinois, NCSA),
Torsten Wilde (Leibniz Supercomputing Centre),
Florent Parent (Calcul Québec),
Susan Coghlan (Argonne National Laboratory),
Natalie Bates (EE HPC WG & ORNL),
Wu-chun Feng (Virginia Tech & Green500),
Erich Strohmaier (LBNL & Top500)

Why a New Methodology?

Our Goals

1. Create a standard method for accurately evaluating energy efficiency at full-system scale
2. Quantify and over time improve large-scale measurement quality
3. Standardize the measurement methodology of rankings like the Top500, Green500, and Green Graph 500

Alternatives

- SPEC Power and Performance Benchmark Methodology
 - High-quality, well specified, but not designed with supercomputer-scale issues in mind
- Green500 Run Rules
 - Easy to measure at scale with coarse-grain accuracy

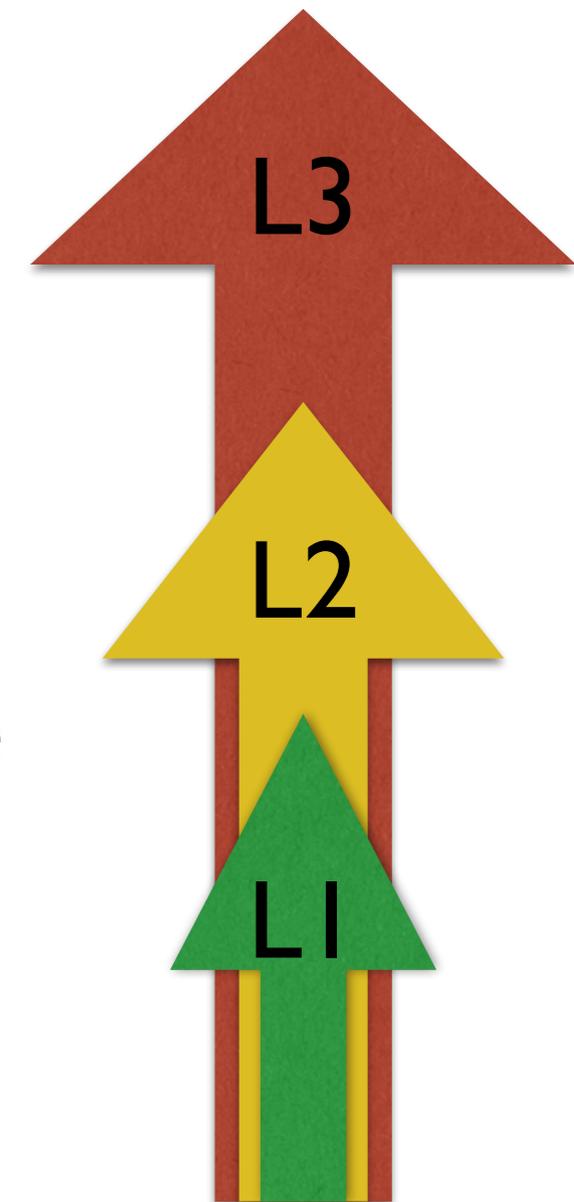
The Methodology

Basic Principles

- Specify the basic requirements to appropriately measure power across a system running a workload
 - No specific workload is required
 - High-Performance Linpack (HPL) is used as an **example**
- Address “gaps” in the previous Green500 and Top500 methodologies
 - What system components to measure
 - What components are system components
 - Scale to varied levels of instrumentation while quantifying the resulting measurement quality

Flexibility: Three+ Quality Levels

- Level 1: Green500 run-rules compatible
- Level 2: Greater accuracy and coverage
- Level 3: Current best accuracy and coverage
- Level 4+: To be determined as new technologies or requirements emerge



Difficulty₇

THE GREEN
500

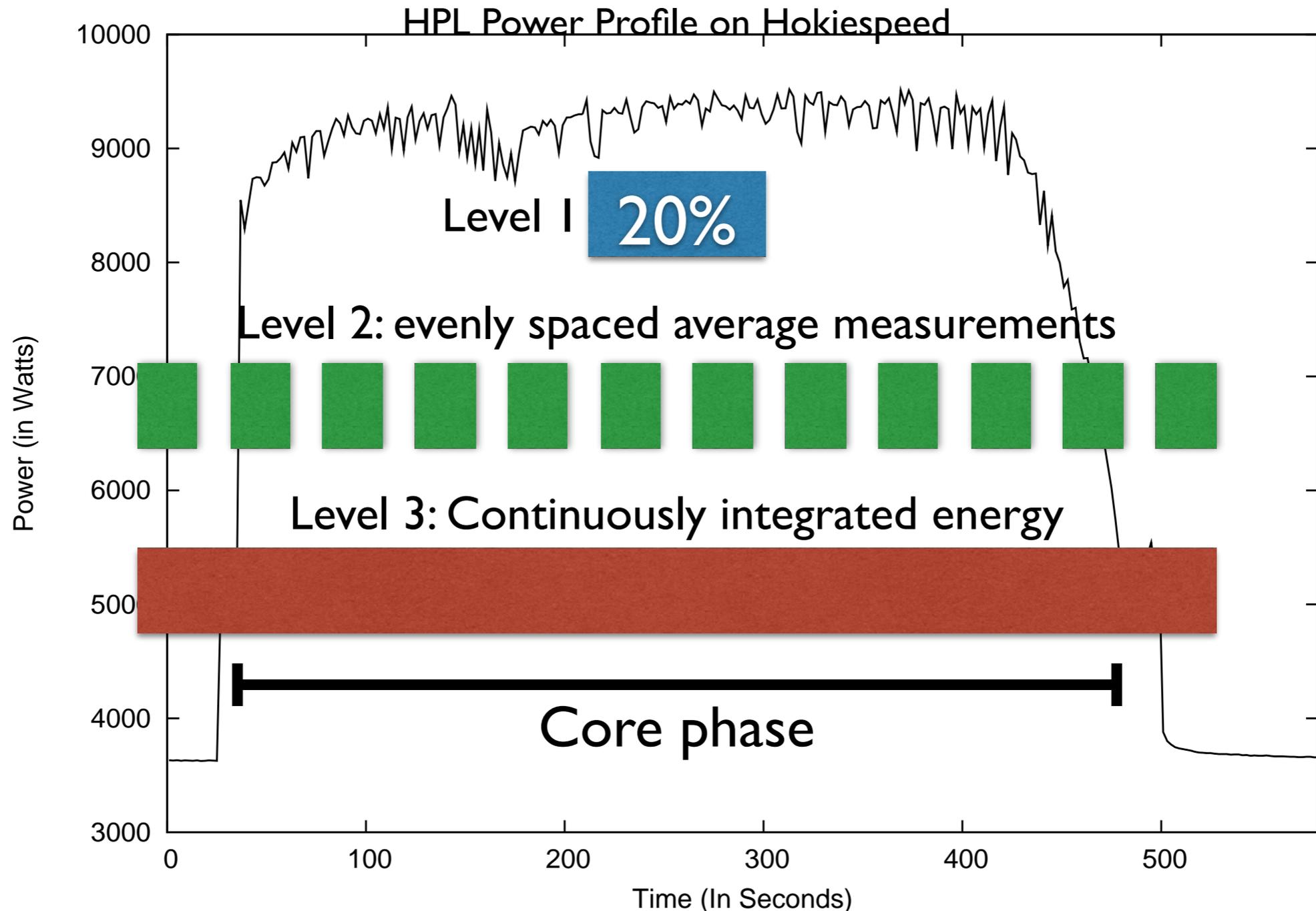
<http://www.green500.org>

Four “Aspects” Define Measurement Quality

The level achieved by a measurement, is the minimum of the levels achieved in each aspect

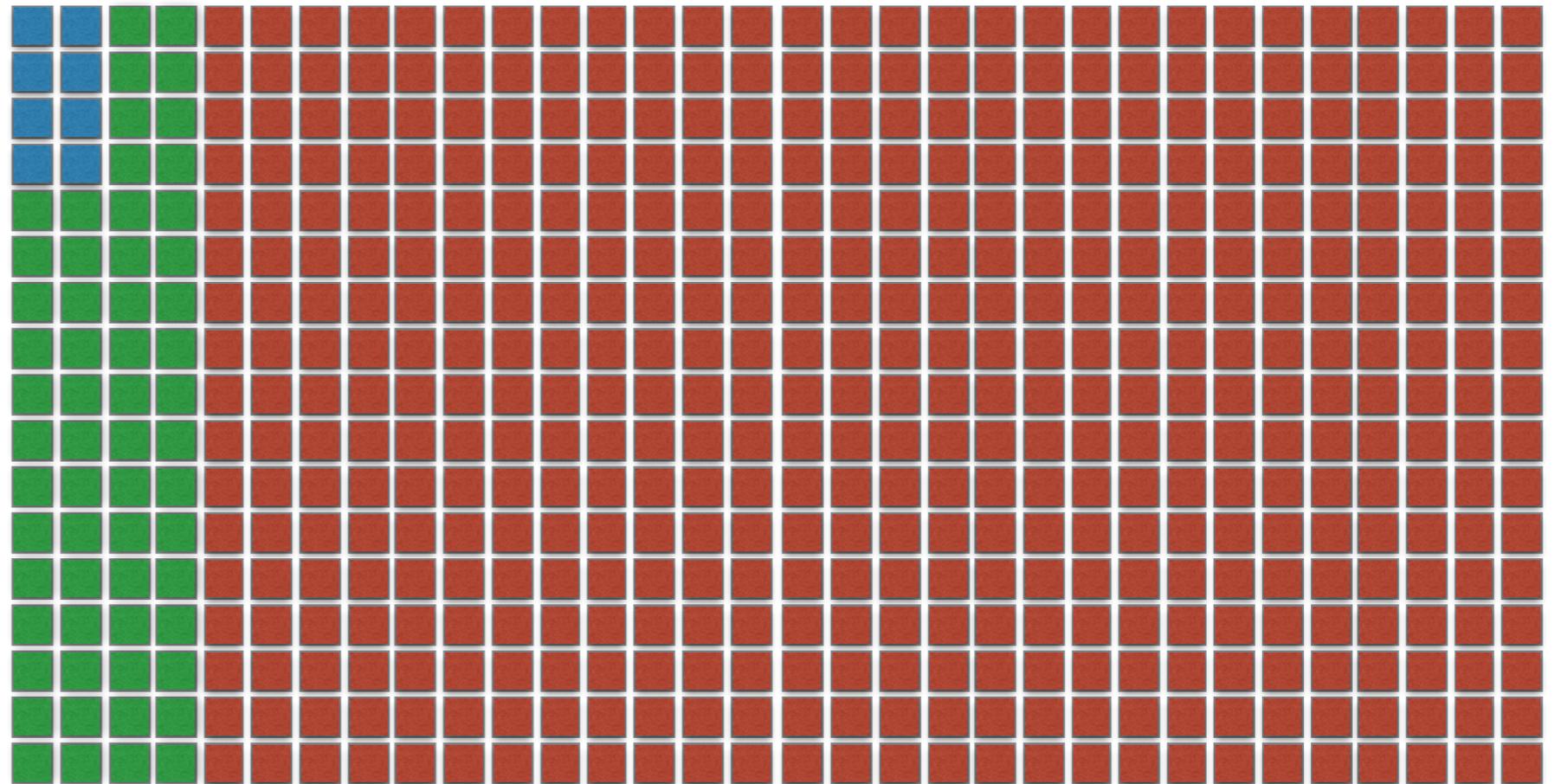
1. Measurement timing and granularity
2. Fraction of system measured
3. Subsystems measured
4. Power measurement location:
 - Before AC->DC conversion or measure conversion loss

Aspect I: Measurement Timing



Aspect 2: Fraction of System Measured

1. 1KW or 1/64th
2. 10KW or 1/8th
3. The whole machine



Aspect 3: Subsystems Measured

- Level 1: Compute nodes only
- Level 2: All participating systems, including networking, storage etc. must be included if used either measured or estimated
- Level 3: All participating subsystems must be measured

Note: Only cooling included within another subsystem is currently required

Case Studies (In Brief)

1. Argonne National Laboratory: Mira
2. Calcul Québec Université Laval: Colosse
3. Leibniz Supercomputing Centre: SuperMUC

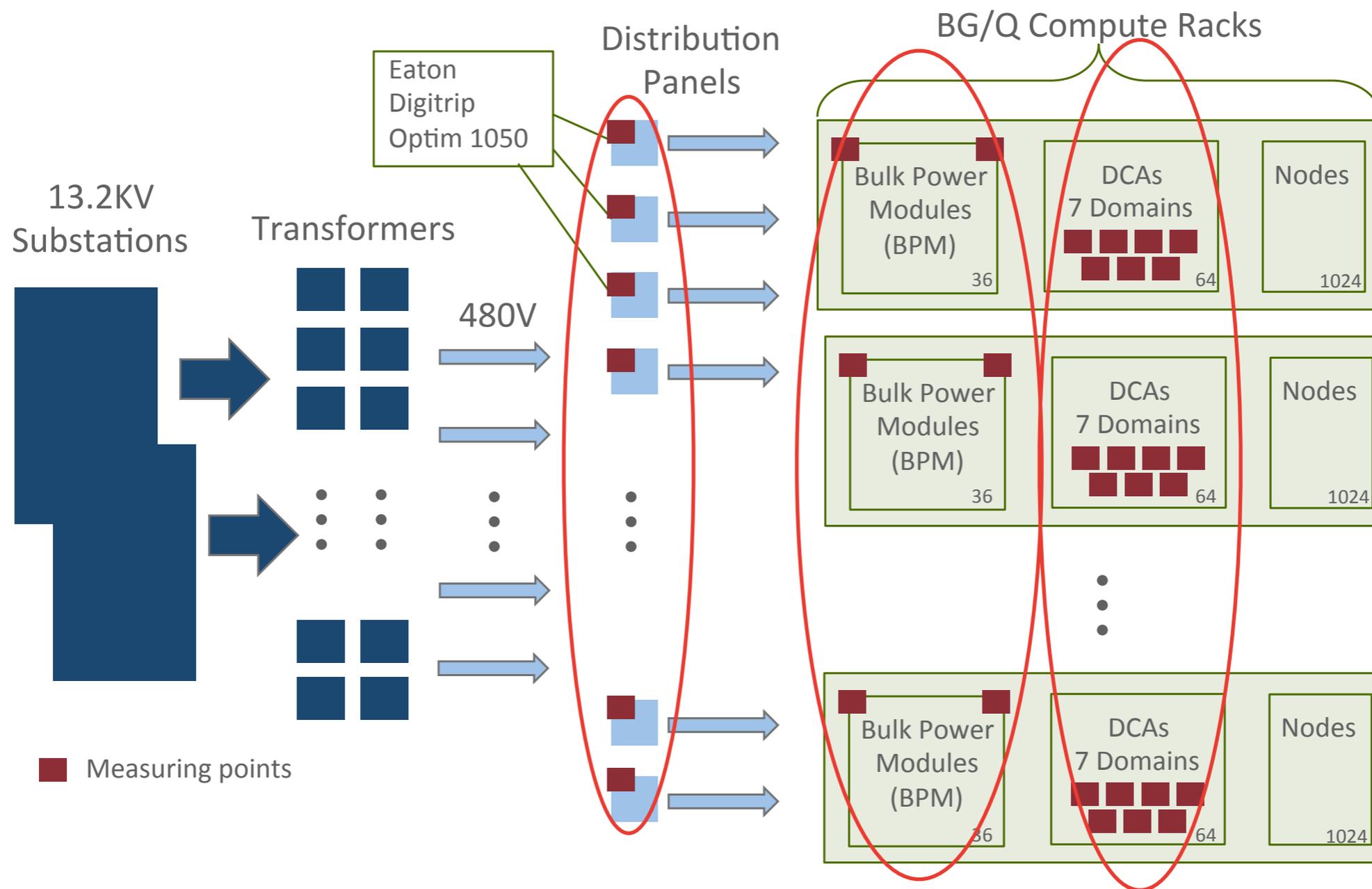
Argonne National Laboratory: Mira

- IBM Blue Gene/Q
- 48 Racks and 48,000 compute nodes
- Evaluated: Levels 1 and 2



Metering and Access

Mira 480V Compute Power Distribution System

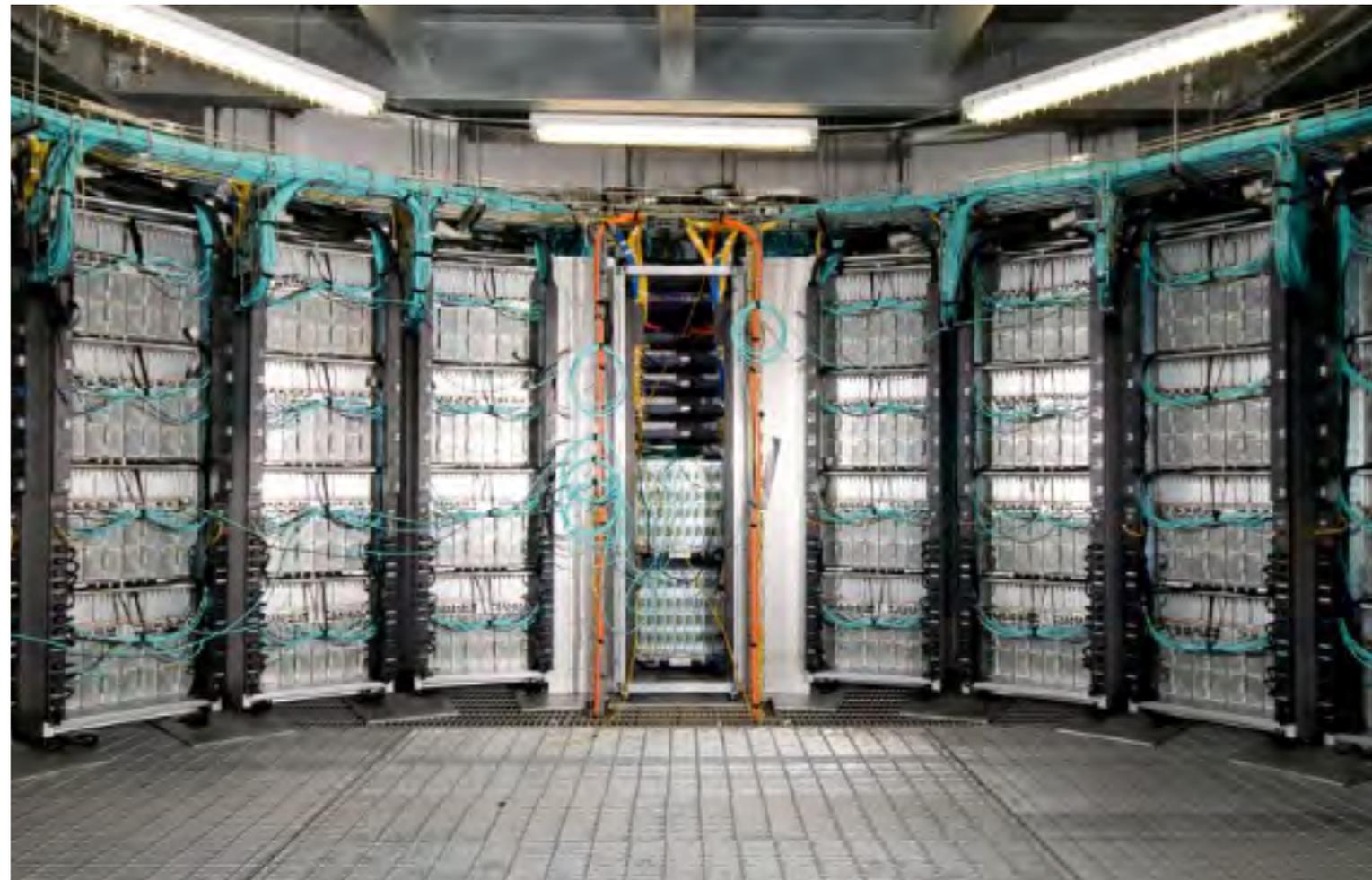


Mira: Lesson

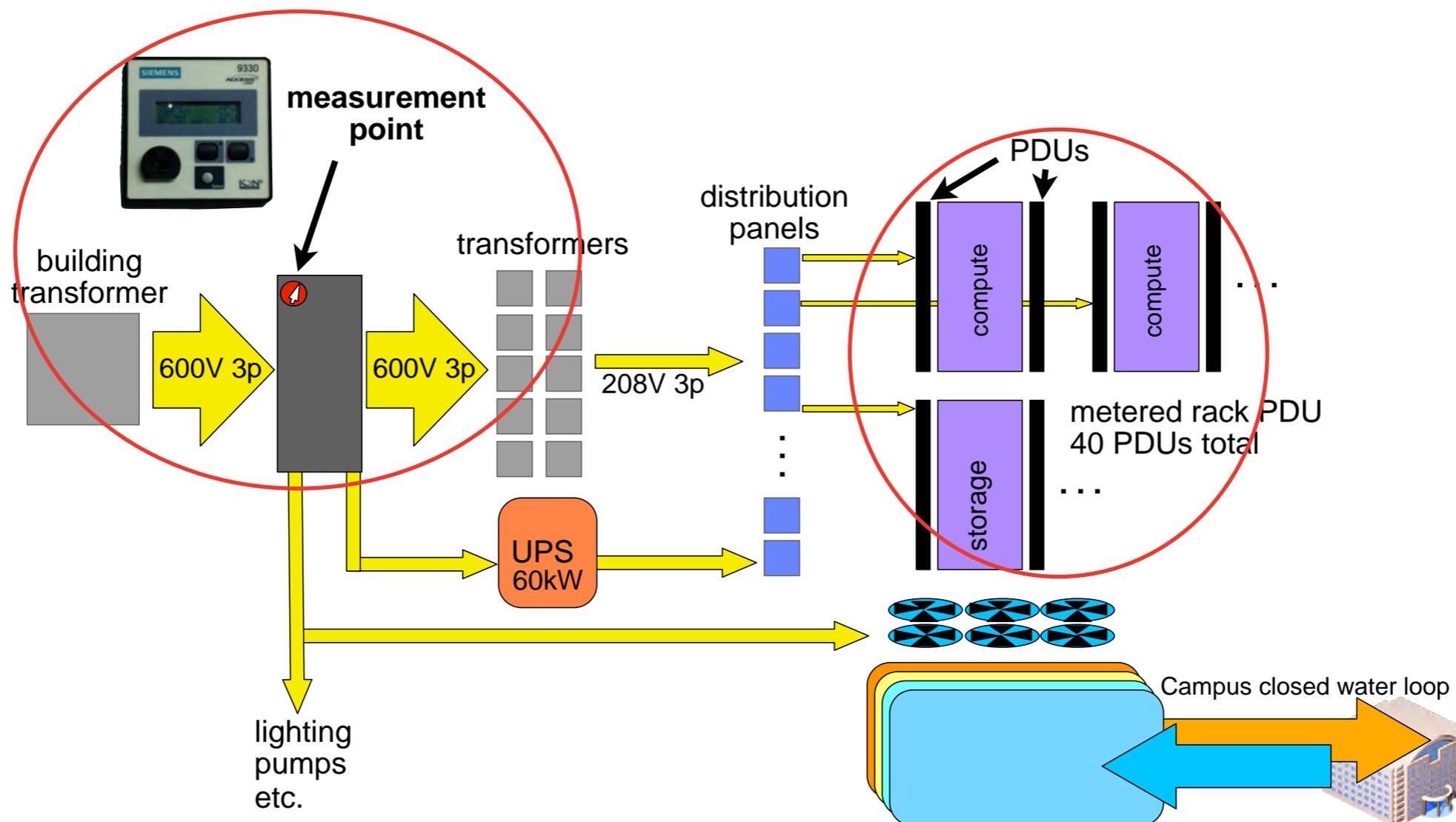
Having the infrastructure is not
enough, it must be accessible to be
useful

Calcul Québec Université Laval:Colosse

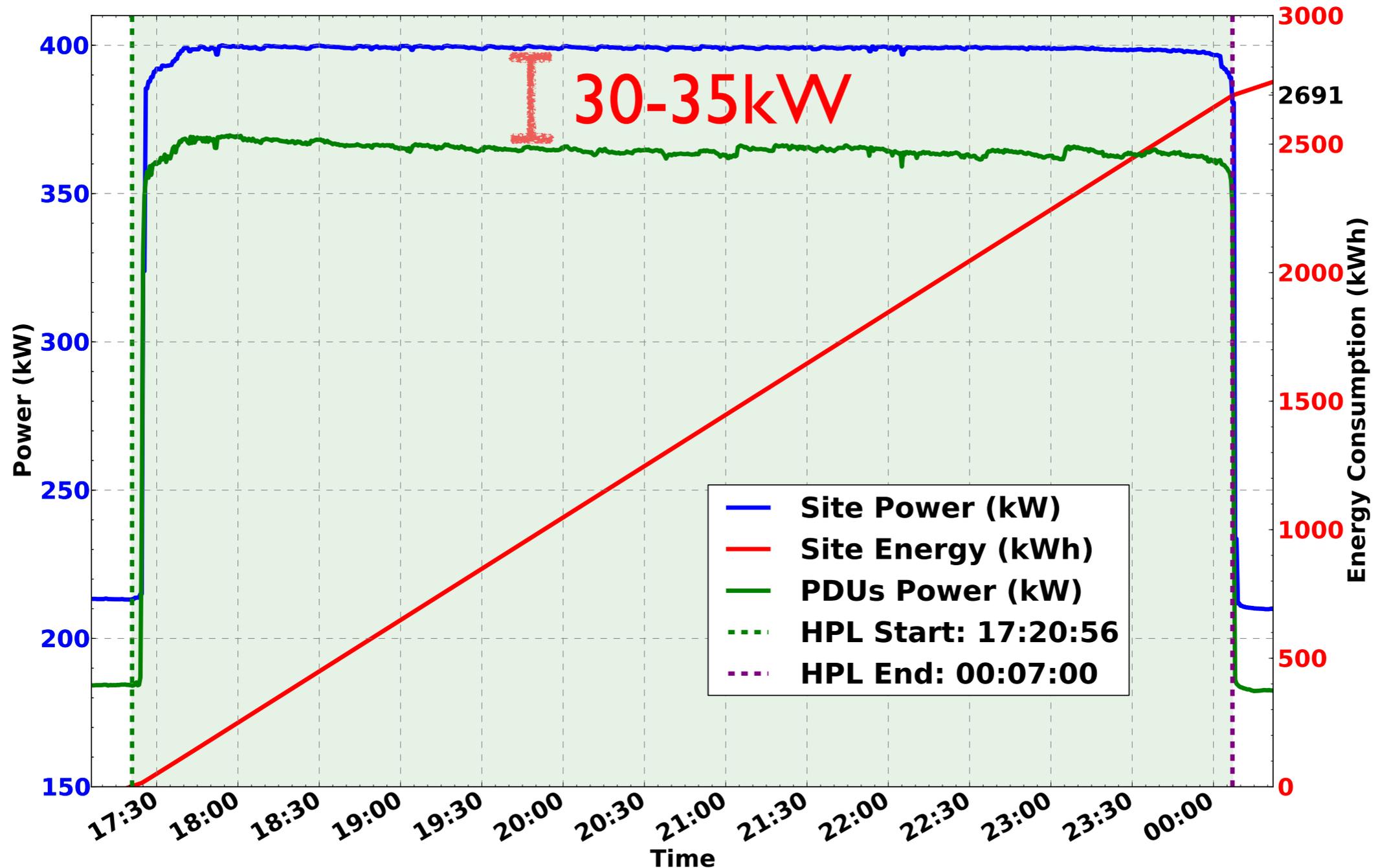
- Sun 6048
- 7680 cores in 960 nodes
- Evaluated: Level 3



Colosse Measurement Infrastructure



Colosse HPL Power: Infrastructural vs. PDU

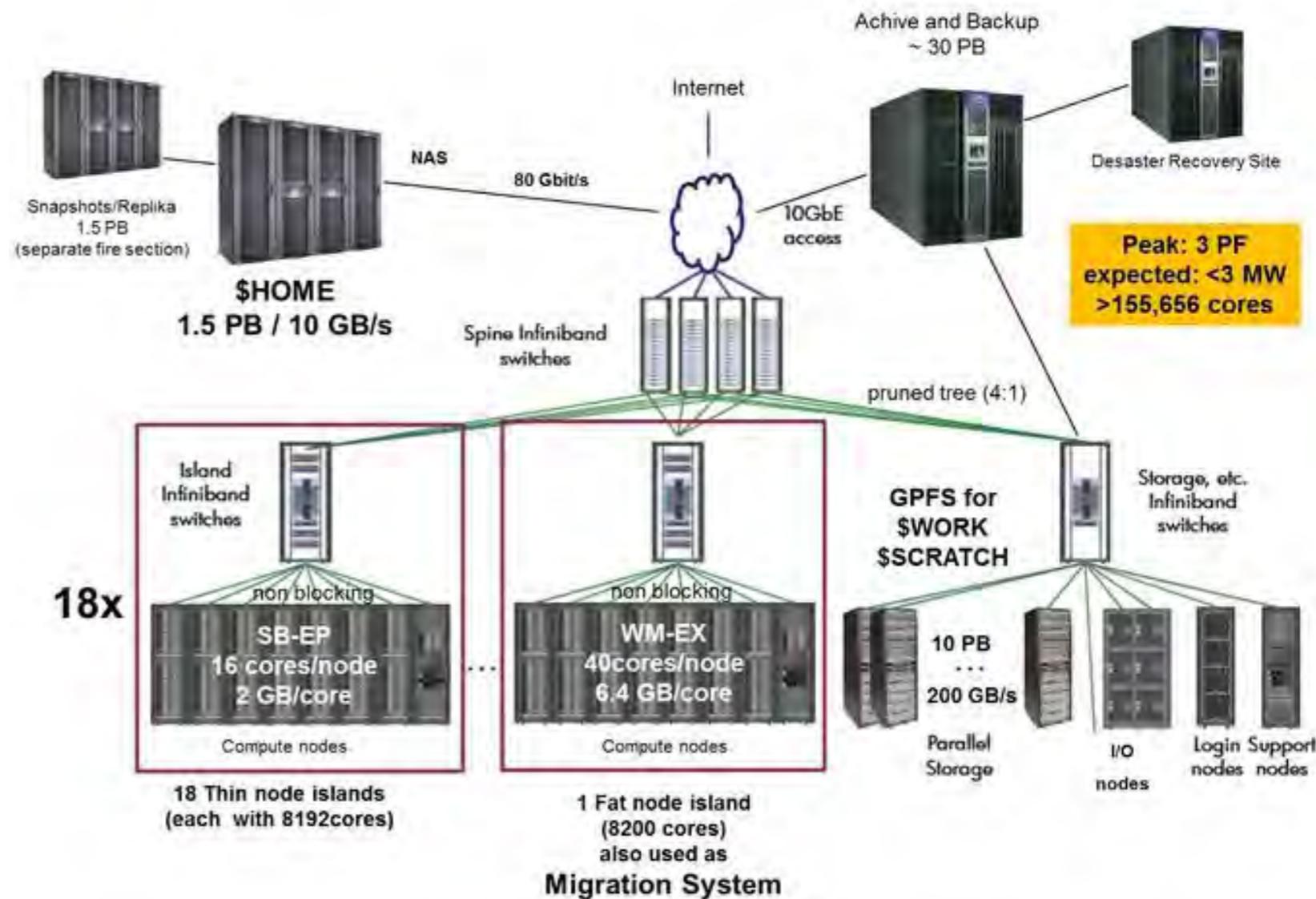


Colosse: Lesson

Large-scale systems are difficult to separate from other infrastructure, prefer high-quality meters at finer granularity

Leibniz Supercomputing Centre: SuperMUC

- IBM System x iDataPlex nodes
- 155,656 processor cores in 9400 compute nodes
- Evaluated: All three levels, standard for Level 3 quality



Results Across Levels: Single Run

Quality Level	Mflops/Watt full run	Efficiency Drop From Level 1
L1 (compute only)	1055	0
L2 (>10kW) (compute and interconnect)	1011	44 (~4%)
L2 (>1/8) (compute and interconnect)	994	61 (~6%)
L3 (compute, interconnect, storage, cooling, power distribution)	887	168 (~16%)

SuperMUC: Lesson

Measurements taken at different
levels are *not* comparable

Conclusions

- We present a higher-quality measurement methodology for use with large-scale HPC
- Even well-instrumented systems easily run into problems measuring a full system
- Comparing measurements at different levels is **unwise**
- High-quality measurements require consideration during system procurement and computing center design